

A method of identifying a relationship between a set of one or more candidate biomolecules and a set of one or more reference biomolecules, the method including inputting to a computer a query set describing the one or more candidate biomolecules; comparing the query set with a target database describing the one or more reference biomolecules wherein the one or more reference biomolecules grouped into one or more buckets and wherein the one or more reference biomolecules of each bucket share a common property; counting a number of matches between each query set and each buckets of the target database; and statistically analyzing the number of matches to each bucket wherein the presence of a statistically significant match identifies a relationship between a the query set and a bucket of the target database.